Alluvial Diagram with ggalluvial::CheatSheet

Introduction

Alluvial Diagrams are a type of **flow diagram** which can be used to visualize frequency distributions over time or frequency tables involving several categorical variables.

Installation

```
install.packages("ggalluvial")
```

Package Load

```
library(ggalluvial)
library(patchwork)
```

What're in the Package?

- Geom/Stat related functions
  - geom_alluvium (GeomStratum) stat_alluvium (StatAlluvium)
  - geom_flow (GeomFlow) stat_flow (StatFlow)
  - geom_lode (GeomLode)
  - geom_stratum (GeomAlluvium) stat_stratum (StatStratum)
- data transform functions
  - to_alluvia
  - to_alluvia_form
  - to_lodes
  - to_lodes_form
  - is_alluvia_form
  - is_alluvial
  - is_alluvial_alluvia
  - is_alluvial_lodes
  - is_lodes_form
- datasets predefined
  - majors
  - vaccinations

Data Structure And Data Used

ggalluvial recognizes two formats of **"alluvial data"**:

- Alluvia (wide) format

```
is_alluvia_form(data, ..., axes = NULL, weight = NULL, logical = TRUE,silent = FALSE)
to_alluvia_form(data, key, value, id, distill = FALSE)
```

- Lodes (long) format

```
is_lodes_form(data, key, value, id, weight = NULL, site = NULL, logical = TRUE,silent = FALSE)
to_lodes_form(data, ..., axes = NULL, key = "x", value = "stratum",id = "alluvium", difftuse = FALSE, discern = FALSE)
```

Example

Using dataset of **Titanic** for illustration.

```
data(Titanic)
dat<-as.data.frame(Titanic)
```

Titanic is an array(tabular) form which stores multiple categorical dimensions such as

- Sex
- Age
- Class
- Survived

For convinience, base::data.frame() was used to convert such an **array** to an acceptable **data.frame**.

```
is_alluvia_form(dat)
```

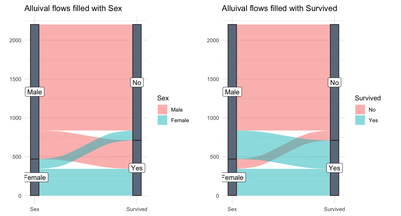
Two Categorical Dimensions

Using ***Sex*** and ***Survived*** variables for illustration:

```
pp1<-ggplot(dat,aes(y = Freq, axis1 = Sex, axis2 = Survived)) +
        geom_alluvium(aes(fill=Sex),width = 1/12)+
        geom_stratum(width = 1/12, fill = "#6D7BBD", color = "#2B1B17") +
        geom_label(stat = "stratum", aes(label = after_stat(stratum))) +
        scale_x_discrete(limits = c("Sex", "Survived"), expand = c(.05, .05)) +
        theme_minimal()+
        ggtitle('Alluival flows filled with Sex') +labs(y='')

pp2<-ggplot(dat,aes(y = Freq, axis1 = Sex, axis2 = Survived)) +
        geom_alluvium(aes(fill=Survived),width = 1/12)+
        geom_stratum(width = 1/12, fill = "#6D7BBD", color = "#2B1B17") +
        geom_label(stat = "stratum", aes(label = after_stat(stratum))) +
        scale_x_discrete(limits = c("Sex", "Survived"), expand = c(.05, .05)) +
        theme_minimal()+
        ggtitle('Alluival flows filled with Survived') +labs(y='')


pp1+pp2
```



Elements of above **Alluvial Diagram**:

- **Axes**: multiple horizontally-distributed bars representing factor variables, such as ***Sex*** and ***Survived***
- **Strata**: vertical divisions of these axes representing these variables' values such as ***Male/Female(Sex)*** and ***yes/no(Survived)***
- **Alluvial flows**: splines connecting vertical subdivisions (**lodes**) within strata of adjacent axes representing subsets or amounts of observations that take the corresponding values of the corresponding variables.

So propotion of strata at each axis and their connections can be visualized from be Alluvial plot.

Alluvial flows can be shown according to **Aesthetic mappings**, in above figure, Sex and Survived were used to "fill" the **flow**.

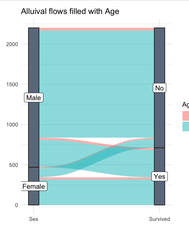
Third Categorical Dimension Added

For the **Titanic** data shown above in Alluvial Diagram, how do we introduce a 3rd variable(dimension)?

Age will be used as the 3rd variable for illustration.

3rd variable as Aesthetic Mapping

```
ggplot(dat,aes(y = Freq, axis1 = Sex, axis2 = Survived)) +
        geom_alluvium(aes(fill=Age),width = 1/12)+
        geom_stratum(width = 1/12, fill = "#6D7BBD", color = "#2B1B17") +
        geom_label(stat = "stratum", aes(label = after_stat(stratum))) +
        scale_x_discrete(limits = c("Sex", "Survived"), expand = c(.05, .05)) +
        theme_minimal()+
        ggtitle('Alluival flows filled with Age') +labs(y='')
```

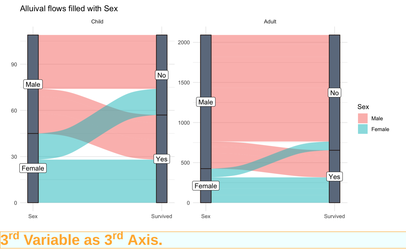


For **Sex** axis, the **lobes** were (from top to bottom):

- Sex(Male)– Age(Child) –Survived(no)
- Sex(Male)– Age(Adult) –Survived(no)
- Sex(Male)– Age(Child) –Survived(yes)
- Sex(Male)– Age(Adult) –Survived(yes)
- Sex(Female)– Age(Child) –Survived(no)
- Sex(Female)– Age(Adult) –Survived(no)
- Sex(Female)– Age(Child) –Survived(yes)
- Sex(Female)– Age(Adult) –Survived(yes)

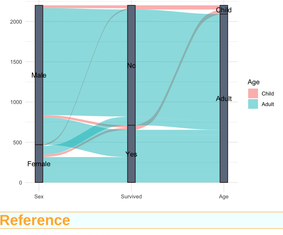
3rd Variable as Facets.

```
ggplot(dat,aes(y = Freq, axis1 = Sex, axis2 = Survived)) +
        geom_alluvium(aes(fill=Sex),width = 1/12)+
        geom_stratum(width = 1/12, fill = "#6D7BBD", color = "#2B1B17") +
        geom_label(stat = "stratum", aes(label = after_stat(stratum))) +
        scale_x_discrete(limits = c("Sex", "Survived"), expand = c(.05, .05)) +
        theme_minimal()+
        ggtitle('Alluival flows filled with Sex') +labs(y='') +
        facet_wrap(~ Age, scales = "free_y") +
        coord_cartesian(clip = 'off')
```



3rd Variable as 3rd Axis.

```
ggplot(dat,aes(y = Freq, axis1 = Sex, axis2 = Survived, axis3=Age)) +
        geom_alluvium(aes(fill=Age),width = 1/12)+
        geom_stratum(width = 1/12, fill = "#6D7BBD", color = "#2B1B17") +
        geom_text(stat = "stratum", aes(label = after_stat(stratum)),size=4) +
        scale_x_discrete(limits = c("Sex", "Survived","Age"), expand = c(.05, .05)) +
        theme_minimal()+
        ggtitle('Alluival flows filled with Age, 3 axes') +labs(y='') +
        coord_cartesian(clip = 'off')
```



Reference

ggalluvial:http://corybrunson.github.io/ggalluvial/